DOCUMENT RESUME

ED 058 309 24 TM 001 024

AUTHOR Dyer, Henry S.; Rosenthal, Elsa
TITLE State Educational Assessment Programs: An Overview.
INSTITUTION ERIC Clearinghouse on Tests, Measurement, and Evaluation, Princeton, N.J.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
REPORT NO TM-R-6
PUB DATE Dec 71
CONTRACT OEC-0-70-3797(519)
NOTE 13p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Community Consultant Programs; Educational Finance; Educational Needs; *Educational Objectives; Evaluation; Evaluation Criteria; Evaluation Techniques; *Program Evaluation; State Action; State Departments of Education; *State Programs; *State Surveys; *Testing Programs
IDENTIFIERS Belmont Project

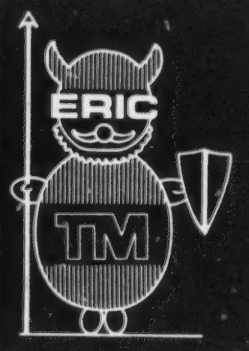
ABSTRACT

This is an edited version of the overview chapter from a report on a 1971 survey of state educational assessment programs. The procedures used in carrying out the survey are described, and a number of major trends in the approach to state assessment are discussed. Finally, some of the more important problems encountered by the states in their assessment efforts are outlined. (AG)

ED058309

TM REPORTS

NUMBER 6



State Educational Assessment Programs
An Overview

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

TM 001 024

ERIC CLEARINGHOUSE ON TESTS, MEASUREMENT, & EVALUATION ■ EDUCATIONAL TESTING SERVICE, PRINCETON, NEW JERSEY 08540
Conducted by Educational Testing Service in Association with Rutgers University Graduate School of Education

This *TM Report* is an edited version of the overview chapter from a report on a 1971 survey of state educational assessment programs. The writing of this overview chapter was sponsored by ERIC/TM.

The full report, *State Educational Assessment Programs*, is published and copyrighted by Educational Testing Service. The report is available from ETS in Princeton, N. J., at a cost of $4.00 per copy.

PREVIOUS TITLES IN THIS SERIES

1. Developing Criterion-Referenced Tests
   ED 041 052
2. Test Bias: A Bibliography
   ED 051 312
3. Ability Grouping: Status, Impact, and Alternatives
   ED 052 260
4. Developing Performance Tests for Classroom Evaluation
   ED 052 259
5. Tests of Basic Learning for Adults: An Annotated Bibliography
   TM 000 987 (ED number not yet available)

The Clearinghouse operates under contract with the U. S. Department of Health, Education and Welfare, Office of Education. Contractors are encouraged to express freely their judgment in professional and technical matters. Points of view expressed within do not necessarily, therefore, represent the opinions or policy of any agency of the United States Government.

December 1971

STATE EDUCATIONAL ASSESSMENT PROGRAMS: AN OVERVIEW

Henry S. Dyer and Elsa Rosenthal

I. INTRODUCTION

In the middle 1960s three events in the national scene had a considerable impact in changing ways of thinking about educational assessment at the state level. The first was the formation in 1964 of the Exploratory Committee on the Assessment of Progress in Education, which eventuated in the National Assessment program now underway.[1] The second event was the enactment of the Elementary and Secondary Education Act of 1965, which included a requirement that school systems assess by objective means the effects on student achievement produced by federally funded programs for the educationally deprived.[2] The third was the publication in 1966 of the Coleman report on *Equality of Educational Opportunity*,[3] which attempted to assess, again in terms of measured pupil achievement, the quality of service the schools were supplying to various segments of the population.

A common element is discernible in all three of these efforts: namely, an insistence that in assessing the performance of the schools major attention must be given to measuring the performance of the children who attend the schools. This approach points up a sharp contrast to the traditional methods of school assessment that had usually appraised the quality of educational programs and services primarily in terms of the quality of school plant and facilities, the paper credentials of professional personnel, the number of dollars expended per pupil, and the like.

Although the three national undertakings mentioned above generated a considerable amount of public controversy, the essential merit of the approach they took has become increasingly clear to educational policy makers at the state level. As a consequence, there has been a growing interest among state authorities in trying to use similar methods for determining what state and local services tend to be most effective in helping students learn.

The states have not been strangers to the concept of measurement in education. Many of them have for a long time sponsored testing programs for a variety of purposes. A survey conducted in 1967, for example, established that there were 74 state testing programs in 42 states, with 18 states offering two or more programs.[4] Most of those programs, however, were at that time intended principally for the guidance of students. Only 17 states were using tests to help evaluate instruction and only 13 to assess student progress. Most of the programs were not in any sense mandatory, nor did any of them provide information about the level and progress of education in the state as a whole. During the last four years there appears to have been a rising demand from state legislators, other state officials, and various public interest groups for this latter kind of information. Accompanied by various political overtones, the question is being asked more and more insistently: "How much and what kinds of measurable pupil learning and development is the state educational tax dollar buying?"

It is against this background that a new survey of state educational assessment programs was initiated in the fall of 1970. The survey has been a joint enterprise involving the Education Commission of the States, Educational Testing Service, and the ERIC Clearinghouse on Tests, Measurement, and Evaluation. The purpose of the survey was simply to find out as much as possible about what the states are planning and doing with regard to statewide educational assessment, what sorts of problems they are encountering in the process, and how they are coping with these problems. It is hoped that the information produced by this survey will help state education authorities achieve a better understanding of the possibilities open to them and the pitfalls to be avoided as they move into the assessment process.

The overall impression one gets from the survey is that state assessment plans and programs are currently in a highly fluid state, with new developments occurring daily. Accordingly, the facts and surmises presented in this report may well be out-of-date within a matter of months. It is for this reason that the entire survey should be viewed only as a snapshot of the situation existing early in the year of 1971. It is for this reason also that we hope this survey will be the first in a series by which, eventually, it will be possible to chart some trends.

In the next section of this overview we shall describe the procedures used in carrying out the survey. In the third section we shall discuss a number of major trends in the approach to state assessment that seem to be emerging. And in the last section we shall take a look at some of the more important problems that the states are encountering in their efforts.

II. THE SURVEY PROCEDURES

The goal of the survey was to obtain detailed information about educational assessment from all 50 states and the District of Columbia. The first step consisted of identifying in each state the two or three persons—usually officials in state education departments—who were most likely to be able to supply the needed information. The Education Commission of the States (ECS) assumed responsibility for assembling the list of state personnel to serve as contacts, for indicating to them the general purposes of the survey, and for enlisting their cooperation. Educational Testing Service (ETS) then assigned 21 persons from the professional staffs of its several field offices to conduct in-depth interviews with the state personnel identified by ECS.

The interviewing took place during the period from the middle of December to the first of March and on the average required about two days in each state. Each interviewer was furnished with an interview guide,[5] but each was also encouraged to go beyond the guide, as might be appropriate, in exploring the specific situation as he found it in the field. Accordingly, there is considerable variation in the nature of the interviewers' reports, and tidy statistics for comparing one state with another are lacking—not only because of the interviewers' differing perceptions of what they heard and saw, but also because of the many different ways in which the states are proceeding and the diverse rates at which they are developing their programs, if any.

There were, however, a number of points covered in practically all the interviews. All interviewers, for instance, inquired into the existence and nature of educational needs assessment programs and into what, if anything, was being done about setting educational goals for the state. They asked whether and to what extent lay citizens had been involved in formulating the goals and whether attempts had been made to translate broad goals into specific and measurable pupil performance objectives. Had advisory or policy commissions assisted in planning, and to what degree had assessment programs gone beyond the planning stage to the implementation of a pilot program or possibly one that was fully operational? Who had initiated the program—the state education department, the legislature, or some other agency inside or outside the state government? Was the control of the program centralized in a state agency, or was it dispersed to the local school districts, or to intermediate units?

Funding was another focus of inquiry. Had the legislature appropriated money especially for the purpose of educational assessment, or had the funds come from the federal government or from regular department budgets?

Technical support for assessment programs was also consistently investigated. Were the universities involved, regional educational laboratories, R and D Centers, private agencies?

Occasionally the states were asked two additional questions: 1) Were their programs being related to and assisted by the Federal-State Joint Task Force on Evaluation (the so-called "Belmont Project")?[6] And 2) Was the assessment program in any way involved with a statewide planning-programming-budgeting system?

Testing programs were examined in some detail. What types of measures, if any, were being used? What educational domains were being explored, and how? Were the measures norm-referenced or criterion-referenced? Were test score data being related to community and school factors? What students were touched by the program at what grade levels? Were all students in the selected grades involved or only a sample? Finally, who would share in the resulting test information? How would it be used? What was the climate in which the programs were conceived? How were the public and the profession responding to the effort? What were the political implications?

The reports submitted by the interviewers were in the form of discursive narratives. Each of these narratives was then summarized and sent back to the state agencies to be checked for accuracy. The summaries were then revised as needed.

The revised summaries, presented state by state, form the major part of the complete survey report: *State Educational Assessment Programs*. (See Reference 5.)

III. MAJOR TRENDS IN APPROACHES TO ASSESSMENT

Although the educational assessment activities of the states are extremely varied, some similarities are immediately evident. One activity, for instance, that is universal is the mounting of educational needs assessment programs. Every state has conducted such a program, or is currently doing so, or is planning to recycle a completed one. The pervasive-

ness of this type of activity is readily explained by the fact that needs assessment is tied to receipt of ESEA, Title III funds, as specified in Section 402 of the act as amended.[7] Another activity involves more than half the states—27 at the present writing—in a joint effort to implement the Belmont Project.* Formulating statewide educational goals is still another task in which many of the states are engaged. In this connection there seems to be increasing recognition that a comprehensive set of agreed-upon goals constitutes the essential defining characteristic of any fully developed educational assessment program—that is, one which can be distinguished from the piecemeal ad hoc testing programs of earlier decades. The way the goal-setting process is being conducted by many states represents one of the distinctly new trends picked up by the survey. We now turn our attention to this development.

The Setting of Statewide Educational Goals

The setting of educational goals by the states has been handled in different ways. Some states, for example, have updated broad goal statements adopted in the past, and they have attempted to translate them into measurable pupil performance objectives for each stage of schooling. A case in point is Colorado, which had adopted a set of educational goals in 1962 but never investigated the extent to which the goals were being achieved. Recently, however, as part of the statewide evaluation project now getting underway there, the Colorado Department of Education brought together a representative group of teachers and subject-matter specialists to specify measurable pupil-performance objectives corresponding to the 1962 goals, and, in a series of workshops at the University of Colorado, to develop tests for assessing progress toward each of the objectives. These tests have subsequently been administered on a pilot basis to students in a sample of schools throughout the state. Other states, not so far along in the goal-setting process, have been faced with the necessity of beginning the exercise de novo.

In addressing this problem, their approaches have varied. Some states are relying solely on professional educators for the establishment of statewide goals. Others, however, are also involving citizens from all walks of life in the exercise. The survey results suggest that the latter approach is becoming increasingly frequent.

---

*The Council of Chief State School Officers and the U.S. Office of Education in 1968 jointly agreed to develop and implement a comprehensive educational evaluation system in an effort to consolidate state reporting of the several federal programs as required by law. The initial meetings took place at the Belmont House in Elkridge, Maryland, and the program has become known as the "Belmont Project." Planned for eventual use in all states, the program presently includes 27 pilot states. Representatives of these states, together with USOE personnel, comprise a Task Force responsible for general development and direction of the project. All states are tied into the project through Evaluation Coordinators appointed by their chiefs.

From all accounts, however, bringing citizens and educators together for the purpose of discussing the ends of education can give rise to a process that is often unexpectedly arduous and time-consuming. The state of California, for example, has been going through this exercise for several years and anticipates that a few more years will be needed before the task can be completed. Its experience is illuminating.

Some time ago the California School Boards Association gathered statements of educational philosophy and goals from virtually every school district in the state. An analysis of the material from some 400 districts resulted in 18 definitions of basic goals. Although these 18 goal statements were given no official sanction by the state education authorities, the activity in and of itself has reportedly influenced state legislation, which now calls for the development of a common state curriculum, modified by local options, and which specifies further that the common curriculum shall be based upon some common set of goals and objectives agreed to in advance.

Concurrently with the work of the California School Boards Association, another group of citizens and educators was also concerning itself with the formulation of educational goals for California. This was the Advisory Committee on Achievement and Evaluation set up by the Education Committee of the California Assembly. After well over a year of hearings, the Advisory Committee recommended to the legislature that a state commission on educational goals and evaluation be established, and during the 1969 regular session a Joint Committee on Educational Goals and Evaluation was given a mandate to tackle the problem.

The Joint Committee, whose members are drawn from the Senate, the Assembly, and the State Board of Education, has appointed still another group of educators and citizens to form an Advisory Committee for Guidelines on Goals. Meantime, working with a staff of consultants, the Joint Committee has decided to require each school district to develop its own goals and objectives based upon the forthcoming Guidelines. Ultimately these local goals are to be added to goals developed by the State Department of Education, by educational specialists, and by citizen advisors. Combined and edited, these goals and objectives will be submitted to the State Board of Education in 1973 together with an evaluation system designed to measure their attainment.

A different example of the apparently inevitable twists and turns that seem to accompany citizen participation in the goal-setting process is to be found in the "Our Schools" program in New Jersey. This program, which got underway in the spring of 1969, is being conducted under the aegis of a broadly representative group known as the Advisory Council on Educational Needs Assessment and is staffed by the Office of Planning in the State Department of Education.

The "Our Schools" program is attempting to answer four questions: 1) What do the citizens of New Jersey think their schools should be doing for the children and

adults of the state? 2) How well are the schools of the state currently doing this job? That is, what are the gaps between goals and results? 3) What can be done in the next three to five years to close the gaps? 4) How can progress toward closing the gaps be measured?

Extensive citizen participation is a basic principle of the program. Two statewide conferences to draw up tentative goals were held in the spring of 1970, each involving about 100 representative laymen, professionals, and students. These were followed during the fall and winter of 1970-71 by 18 regional conferences, involving varying numbers of laymen and professionals, to rework the goals and help collect opinions on priorities. The outcomes of these regional conferences will be supplemented by additional conferences at the local district level and by a statewide poll of citizen opinions concerning public education. In the fall of 1971, the data generated by all this activity will be fed to a final statewide conference of about 300 persons who will attempt a final ordering of educational priorities for presentation to the State Board of Education. The Board will then have the responsibility of determining what the educational goals for the state as a whole are to be.

This mingling of laymen and professionals in the several states has occasioned a search for ways to do justice to large numbers of people and points of view and, at the same time, achieve a workable consensus within practical time limits. The survey reveals that some state educational agencies now plan to train their staffs in the use of the Delphi technique,[8] a process that may prove particularly useful in the goals-setting process. The Delphi technique was originally conceived as a way to obtain the opinions of experts without necessarily bringing them together face to face. The experts are consulted individually, as a rule by a series of questionnaires. Although there have been a number of adaptations, the general idea has been to prepare successive rounds of questions that elicit progressively more carefully considered group opinions. Experimentation has revealed that the process is able to produce a satisfactory degree of convergence of opinion.[9] To our knowledge, however, it has not yet been used with the very large numbers of persons and viewpoints such as those encountered, for instance, in the "Our Schools" program in New Jersey. If the trend toward community deliberation on state policy matters continues, there will need to be further adaptations of the Delphi technique in large-scale settings.

Assessment and Management Information Systems

In an earlier time, accounting systems in education were usually called upon for a fairly simple attesting that the public funds for education had been honestly administered. Such systems are now being asked increasingly to display relationships between the expenditures for school programs of various kinds and the benefits accruing from those programs in terms of student performance. As a result there is a notable trend in many states to apply to the management of the educational enterprise the principles of cost-benefit analysis embodied in some form of planning-programming-budgeting system (PPBS) and to tie statewide educational assessment into such a system.[10]

Although progress toward the actual implementation of PPBS has been slow, this is not for want of enthusiasm among its proponents. The plain fact, however, is that many questions must still find answers before complete systems can be designed and confidently applied. For there is still much to be learned about how to isolate the costs of educational programs and about the analytical techniques for relating benefits to costs. Many state education departments are therefore planning to have their staffs trained in the skills requisite to developing and operating PPBS.

In New York State, for example, an adaptation of PPBS, Program Analysis and Review (PAR), is currently used by the State Education Department to help identify program problems, the main applications being the state's ESEA programs. In the future the Department plans to use information from its Basic Educational Data System (BEDS) in the PAR system to evaluate ESEA projects in terms of an input-process-output paradigm.

California has similarly been developing PPBS for several years. The system has already been pilot tested and subsequently revised and retested. Although PPBS is not yet mandated for the entire state school system, reports are that it is likely to be authorized by the legislature and be fully operational by 1973-74.

Hawaii's legislature has recently called upon the State Department of Education to undertake the same kind of effort, since it is eager for data on educational results and is expecting that the new system will furnish the desired information on how well education in the state is faring relative to the amount of money being spent on it. The Department plans to feed into the system data from its well-established state testing program.

The Federal-State Joint Task Force on Evaluation (Belmont Project) may be having a not unrelated impact on the development of state educational management control systems. As noted before, 27 states are now participants in the project's many activities, which at present also include the development of a Management Assessment System for state education agencies and its testing in a few states. It is possible that the kind of thinking and training required for this and related Belmont activities may have a spillover effect on developing rationales and methodologies for statewide assessment systems.

The Belmont group is not only concerned with building instruments for collecting a broad range of information on the nature, cost, and effectiveness of many kinds of educational programs in school districts; it is also concerned, perhaps more importantly, with the development of methods for training state and local personnel in the use of these instruments. As a consequence, Belmont may be seen as a comprehensive effort to bring into being an information system that can possibly have just as much usefulness in the management of state and local educational programs as it may have for federally supported programs.

Assessment and Statewide Testing Programs

Although educational assessment, properly viewed, involves a good deal more than statewide testing programs, testing seems, nevertheless, to be looming larger and larger in the plans for assessment. In fact, many of the authorizations from legislatures are principally for the assessment of education by tests. That is, there is a mounting legislative pressure for documenting the products of the educational process by statewide testing programs. Some states have already set in motion widely ranging programs of tests (Pennsylvania and Michigan being notable examples), and others report themselves to be at the point of doing so (among them Colorado and Delaware). Some states are starting with rather narrow content coverage, but are planning for massive programs later on (Florida and Georgia, for example).

The content of most current state testing programs—whether mandated or unmandated by legislative bodies—is often less surprising than it is significant. The states engaged in some form of assessment-by-testing are mainly concerned with how well their educational systems are succeeding in imparting basic skills. Only a relatively few go beyond the three Rs to get information on how education is affecting student values and attitudes. Arizona, for example, received a mandate for the Arizona State Third Grade Reading Achievement Program, to begin this year. Although the specific objectives of the program are not yet available, strong effort will apparently be made to provide background data to lend depth and perspective in interpreting test scores.

As another example, recent legislation in Michigan calls for measures of the basic skills at grades 4 and 7. This program, which is now in its second year, covers verbal analogies, reading, English (mechanics of written English), and mathematics. In the first year, only average scores by school and school district were reported, since the tests were consciously designed to be too short to yield adequately reliable scores on individual students. This approach, however, was changed for the 1970-71 administrations. Tests are now of conventional length to provide the schools with information concerning the achievement of individuals. Although the major stress here has been on the academic areas, the Michigan program has also given some attention to assessing the influence of schooling on student aspirations.

California, which has a history of mandated testing programs going back to 1961, is another instance where testing of the basic skills has been strongly emphasized. In 1965, the Miller-Unruh Basic Reading Act created an obligatory testing program in reading for the primary grades. This concentration on the basics has been further reinforced by a recent legislative requirement for the adoption of minimum academic standards for certain grades and the selection of tests to be used statewide in evaluating the attainment of these standards.

Delaware is one of the states that is starting small. It is currently testing achievement and mental ability in all schools, but at the fifth grade only. It is looking ahead, however, to a program that will include all students in all grades, K through 12, in all schools, public and private. Program development in other states is following a similar pattern. Florida, for example, is presently concerned with measuring only achievement in reading, but under legislative prodding is also planning a most ambitious program that will sample students in kindergarten through grade 12 in all the basic subjects.

Assessment of Noncognitive Development

Although the principal intent of most state testing programs is to get a reading on the cognitive development of students, a few states make a point of stressing additionally the importance of personal-social development as an outcome of the educational process. Thus, the idea that education is to be construed simply as a process for inculcating the fundamental cognitive skills no longer totally dominates educational thought and practice.

In recognition of the importance to the student and society of noncognitive development, Pennsylvania includes in its targets for quality education a number of attitudes and noncognitive abilities that it wishes its public schools to nurture. Consequently, the state educational agency has produced instruments to gauge how extensively schools are affecting such significant aspects of human life as self-concept, understanding of others, responsible citizenship, health habits, creativity, the acquisition of salable skills, the understanding of human accomplishments, readiness for change, and students' attitudes toward their schools. Michigan, too, has included in its testing program the measurement of three types of student attitudes: namely, attitude toward learning, attitude toward academic achievement, and attitude toward self. Nebraska is now planning to create an assessment program which, in its first stage, will be concerned *only* with nonacademic objectives.

Measuring the Influences on Learning

A fifth trend, and a significant advance in mounting state testing programs, is the commitment on the part of a number of states to assessing the outcomes of education only after accounting for the effects of community and home environment, of teachers and school programs, and of school facilities and financial resources. To judge from the planning reported in the survey, this is a development in the assessment process that presumably will grow in importance, especially if the Belmont Project continues to expand its services and refine its battery of instruments. For example, the most recent plans of the Belmont group are "to demonstrate now that the System can provide meaningful inputs to the State and/or local educational agencies to assist them in the performance of their basic program functions. This can be partially accomplished through development of a model for a State Data Analysis Plan.

Such a model would be designed to indicate the potential uses of Belmont System data in relation to existing state and local data resources and would tie these together as input to the continuing program evaluation required at both State and local levels to meet the information and decision-making needs of program managers at these levels. . ." (See Reference 6; p. 28.)

The Belmont group expects to begin this year to study total state assessment needs. In fact, some of the group's instruments, now being developed, may be of direct service to any state wishing to assess the influence on its schools of input and process factors. Questionnaires have been constructed to elicit information on organization patterns in schools, the training of personnel, programs and services, condition of school facilities, size and location of school, nature and size of staff, and the like. Other instruments supply information that can provide the basis for evaluating program effectiveness, as, for example, data on classroom facilities, classroom organization, programs of instruction, teacher background, and pupil's grade, age, sex, absences, background characteristics, academic program participation, behavior, and performance.

Among current statewide programs, Michigan's, for example, relates all achievement measures to student and school characteristics. Each student anonymously supplies information from which socioeconomic status and aspiration scales are derived. Records maintained in the State Department of Education provide school and district information such as teacher/pupil ratio, financial resources per student, average teacher experience, and location by type of community. Similarly, the program in Pennsylvania attempts to measure input variables of three major types, which include 8 having to do with the student's background, 4 having to do with the community in which the school is located, and 27 that have to do with school staff characteristics. Community conditions are derived from a Student Information Form. Norms have been developed by the Pennsylvania Educational Quality Assessment, so that school districts can compare pupil achievements, taking into account socioeconomic and other differences in pupils, schools, and communities.

These comprehensive approaches to the assessment of the educational process, school by school, are still relatively rare. However, more and more states appear to be getting interested in the possibility of going in the same direction.

Influence of the National Assessment Model

The survey reveals that, as states tool up for assessment, they are considering whether to use some kind of sampling approach—that is, to obtain information from a relatively small but representative group of students located in representative regions and types of communities in the state—or to use an "every-pupil" approach. Settling the issue often appears to depend on how the purpose of statewide assessment is locally perceived.

If the state wishes principally to supply its decision-makers with satisfactory information about the level and progress of the state's educational system as a whole, the sampling approach is regarded as sufficient. In this connection, the survey reveals a rather pervasive influence of the National Assessment model on state assessment designs. This model is based on matrix sampling techniques and randomization in the packaging of test exercises. Under this strategy, only a few pupils in each school or school district try a few test items drawn from very large pools of items.[11] The model is reportedly attractive because it does not subject any pupil to many hours of testing, while at the same time it provides a large quantity of information on what various segments of the student population are learning during the school years. Colorado and Florida are two states whose plans are based on this kind of sampling approach.

If, on the other hand, the state wishes to couple management-oriented results with information that can be returned to each school for self-appraisal and for the guidance of students, then the every-pupil approach is clearly the appropriate alternative. Georgia's plans at present envision this approach.

These, of course, are not the only possibilities open to the states. There are plans and programs that adopt the "whole-test" approach while testing only a sample of the children in selected grades at any given time.

The evidence is not yet clear enough for a prediction of which sampling patterns will ultimately predominate. As programs move past discussion-and-drawing-board stages, future surveys should illuminate further and document the various conditions and considerations that influence choices.

The Control of Assessment Programs

The control of state educational assessment programs follows several patterns. In some states there is a strong tendency toward the centralization of control in the state department of education. In others the tendency is to vest much of the control in the local school districts. In still others, there is a kind of balanced tension between the two tendencies. Nevertheless, the results of the survey suggest that, insofar as testing is a component of assessment, there may be a slight trend toward more centralized control of the assessment process, even in those cases where participation in the program is optional with the local education authorities. In such cases the state authority assumes responsibility for specifying the purpose, content, and target populations of the programs but the local districts may be left free to accept or reject the state's services. At the same time, however, there is a noticeable if small increase in programs whose results are aggregated and analyzed for the entire state and reported by a central agency to legislatures or to state boards of education as well as back to the administrators of the local school districts. This naturally occurs where legislation so stipulates.

The survey also indicates that where some form of centralized operating control exists, the state department of education is not necessarily the agency that exercises it. Indeed, the control may be based in the education department of a state university or, as is the case currently in Texas, in regional centers that have been established by law but which work largely independently of the state department of education.

Thus, local programs of assessment and local options to participate in centralized programs continue unchanged as typical manifestations of the folkways of American education. Yet the survey gives some salience to procedures that begin to combine, in novel and even ingenious ways, the two approaches to control. That is, as the states feel constrained to renew or to rationalize their educational systems, some have adopted models to permit both maximum feasible local autonomy *and* the exercise of state leadership in improving local educational processes. An interesting instance is the Vermont Design for Education. The emphasis here is on the *state's* requiring an extraordinary degree of *local* involvement in educational planning. In effect, Vermont has required each locality to build its own locally created design for education and has also required full citizen involvement in setting goals and priorities. The Vermont Design was created by the state education agency.[12] Its purpose, however, was not to impose programs, but to stimulate vision, discussion, and creativity. This "conversation-piece" model also includes state-developed instruments that the districts are free to use if they wish—or to adapt or reject in favor of locally devised tests and other measures. The state agency also stands ready to offer assistance when the locality is in need of technical expertise. A representative of the Vermont State Department of Education, for example, sits in on community meetings as a source of immediate technical assistance and information. Hence, although there is direct influence, there are no constraints on the form and shape of local programs. The central agency's effect is to lead autonomous localities in the direction of self-determined innovation. The state commissioner will receive formal reports of the resultant programs, but they will not be publicized.

IV. SOME EMERGING PROBLEMS

Embedded in all this state assessment activity we detect a variety of problems emerging which, in our view, will need more attention than they have generally been getting if much of the planning now underway is not to be frustrated. These problems have to do largely with the strategies and tactics by which viable programs of assessment are to be brought into being and maintained. The problems fall into four categories: 1) lack of communication and coordination, 2) the relation of assessment data to financial incentives, 3) the handling of sensitive data, and 4) confusion and conflict about goals.

The Problem of Coordination and Communication

In some states a number of different groups appear to be going their separate ways in moving toward the design of some sort of educational assessment program. These disparate groups may include legislative committees, citizen committees (self-appointed or governor-appointed), state boards of education, state departments of education, and even different segments of the bureaucracy within a state department. The absence of any serious effort to coordinate the efforts of these several groups or to open up lines of communication among them can generate conflict and confusion which threaten to neutralize the entire enterprise.

There is, for example, the recent case of a legislature that adopted two conflicting statutes whereby some of the well-laid plans for one statewide testing program were effectively nullified by the legal specifications for a second program.

In another state three programs appear to be moving independently along nonconvergent parallel lines toward the same ultimate objective. One program under the control of one branch of the department of education is trying to develop a statewide consensus on educational goals; another under the control of a committee of the legislature is trying to develop a state-aid system that will include a requirement that each local school district devise its own appropriate goals; and a third under the control of another branch of the department of education is looking toward a statewide evaluation program based upon a set of goals not yet determined.

In yet another state at least four different programs, each under separate auspices and each separately staffed, are in various stages of development. One of these is being developed by a governor-appointed commission which is looking into school financing and assessment programs that might be devised to rationalize the process. A second, located in one of the divisions of the state department of education, has been providing, on an optional basis for a number of years, a battery of tests and other measures whereby a school system may, if it wishes, assess the effectiveness of its instructional programs. A third, operated by another division of the same department, administers a statewide testing program on a required basis to all elementary schools in the state and provides its own advisory services to help school personnel use the results to evaluate educational progress by comparison with state norms.

Finally, still another branch of the department, using different data, has been working for several years on checking out the feasibility of an input-output model for measuring school effectiveness.

Diversity in the efforts to build an educational assessment system for a state is probably inevitable as a consequence of professional and political rivalries among the several groups concerned. It can be argued that such diversity in some amount is desirable in that it may help to ensure that a system best adapted to the state's needs will eventually emerge.

On the other hand, when fragmentation of the planning activities becomes so extreme that there is little if any communication among the planners, the whole effort can be counterproductive in at least two ways. It can create so much confusion in the local school districts that they will tend to sabotage any and all assessment programs that may be forthcoming. And it can result in so much duplication of effort as to be wasteful of time, money, and the technical expertise that is still extremely scarce.

Accordingly, if state educational assessment is to fulfill its very real promise as an instrument for helping educational systems upgrade the quality of their services, it would appear that means must be found for exchanging ideas about what a sound assessment program in a given state might be and for encouraging cooperation among those involved in the development of programs.

The Relation of Assessment Data to Financial Incentives

Another problem beginning to crop up where statewide assessment programs are actually underway has to do with the manner in which the results will be used in allocating state funds to local school districts. One can put the problem in the form of four questions:

1) Does one use the funds to reward the districts that show up high on the indicators?
2) Does one withhold the funds to punish the districts that show up low on the indicators?
3) Does one use the funds to help upgrade the districts that show up low on the indicators and thereby withhold funds from those that show up high?
4) Or can one find a way to allocate the funds so that all districts will have an incentive for constantly improving the quality of their schools?

These are agonizing questions that have apparently not been adequately thought through. For example, one state is now using reading test scores in a formula for determining the specific sums of money that will be allocated to school districts to provide reading specialist teachers. Depending on the progress of the students, the school can suddenly find itself without funds for specialized assistance because it has previously been successful in improving reading levels.

In another state—where there is similar legislation—funds are being awarded to schools that rank lowest on common measures. Some school principals who are serious about their responsibilities are beginning to talk of deliberately over-speeding test administrations so that school performance as measured by the tests will *not* come up to the mark. Their reasoning apparently is that if failure is to be rewarded, then it is folly to be successful.

Sound answers must be found to these questions. If they are not, the whole assessment enterprise runs the risk of provoking the outrage of both the public and the professional educators.

The Handling of Sensitive Data

One particularly troubling problem beginning to surface has to do with the confidentiality of information supplied by pupils, teachers, and others who may be involved in some aspect of the assessment process. The question arises in the first instance in connection with the release of achievement test scores of individual pupils and the averages of such scores, class by class, or school by school, or even, in some cases, district by district. The fear is that data of this sort will be misinterpreted by the public and be used to make unwarranted and invidious comparisons.

The problem is further exacerbated when pupils and/or their teachers are asked to supply information about their ethnicity, their economic and social backgrounds, their behavior tendencies, and their social attitudes. Hard questions are raised not only concerning the propriety of using such information once it is in hand, but also concerning the possible deleterious effects on children of merely asking for such information in the first place. It is argued, with some cogency, for instance, that to ask a child from a broken home "Who acts as your father?" can be psychologically damaging to the child; it can also be regarded as invading privacy.

Furthermore, there is always the doubt whether the responses to such questions can be taken at face value as a true representation of the child's home conditions. Similarly, in respect to questions about attitudes, the doubt is always present whether the respondent may be "faking good" or "faking bad" and not representing his true feelings about himself and others.

As a consequence, any comprehensive assessment program that attempts to secure data on the many interacting variables bearing upon the multiple outcomes of the educational process is confronted with a serious dilemma. Unless the kinds of sensitive data suggested above become available, any assessment of what schools are doing to and for students will be less than complete and very likely misleading. On the other hand, the ethical and practical difficulties in collecting such data are very real difficulties that are not easily overcome.

Recently, for example, some schools involved in a state testing program refused to return the students' answer sheets on the ground that the responses they contained might be used to penalize the individual student because of his background or possibly to impugn the reputation of his ethnic group. And this reaction occurred despite the fact

that the information was gathered in a manner that guaranteed the anonymity of the suppliers thereof and despite the announced intent to use the information only for the purpose of assessing the overall impact of educational programs on each of several target populations of students. In short, even though the state authority may be doing its best to protect the integrity of the data required for giving the public a reasonably accurate picture of the educational benefits its tax dollars are buying, the public in turn is often so dubious of the credibility of the state authority in these matters that efforts to develop sound assessment procedures are in danger of reaching an impasse.

Some attempts have been made to circumvent the sensitivity-of-data problem by relying on various types of "social indicators." This is done by using existing data collections—for example, federal, state, and local statistical reports on community economic status, health, juvenile delinquency rates, the use of public libraries, concert halls, museums, and the like. Each such indicator is presumed to be capable of giving some indirect information relative to the overall impact of schooling on children. However, the difficulty with these kinds of indicators of school effects is well-known and far from being dispelled.[13] The difficulty inheres in their very indirectness, in the fact that the level of such indices is determined by many social and community factors beyond the reach of the schools.[14] Hence, they are highly vulnerable to misinterpretation.

Confusions and Conflicts about Goals

In the various efforts to formulate meaningful goals upon which to build assessment programs, there appears to be a considerable amount of confusion between the ends and means of education, between process and product, between inputs and outputs, and between pupil performance objectives, staff objectives, and system objectives. This sort of confusion pervades not only public discussions of educational goals; it appears to be just as rife in the deliberations of the professional educators themselves.

The following list of abbreviated goal statements is not unrepresentative of the kind of mix such discussions frequently produce:

- . . . To help students become effective participants in society
- . . . To increase the ratio of guidance counselors to pupils
- . . . To ensure that students acquire sound health habits
- . . . To ensure that all students are capable of reading "at grade level"
- . . . To reward teaching and administrative personnel in accordance with the degree to which they produce learning in students
- . . . To reduce class size by increasing the ratio of teachers to pupils
- . . . To provide more effective in-service training for school personnel
- . . . To ensure that every student shall have acquired a marketable skill by the time he or she graduates from high school
- . . . To stimulate community involvement in the work of the schools
- . . . To reduce the student dropout rate
- . . . To modernize and enlarge school facilities
- . . . To give students a sense of their worth as human beings
- . . . To keep school budgets as low as possible consistent with sound education
- . . . To sensitize teachers to the individual learning needs of the children they teach
- . . . To bring the results of research to bear on the actual operations of the schools
- . . . To promote better understanding among ethnic, racial, and economic groups

The difficulty with such an indiscriminate collection lies in the fact that the individual goal statements, however worthy in themselves, are so diverse in type that there is no way to compare them with one another and thereby arrive at priorities among them. Some attempts have been made to get around this difficulty by sorting the goals into homogeneous categories of objectives, such as societal objectives, pupil performance objectives, process objectives, staff requirement objectives, financial objectives, and the like. Even so, however, the vexing problem of how to work out the probable *interrelationships* among the several categories has seldom been addressed in any explicit way. Nor, despite the efforts of system analysts to develop the necessary conceptual schemes and procedures for rationalizing the relationships, does there appear to be much inclination among educational policy makers and practitioners to come to grips with the problem.

One reason for this state of affairs seems to lie in the very real complexity of the goal-making process. It is no mean task to sort out, even in rough fashion, the several types of goals, to make them operational in terms of defining measures, and to visualize the possible relationships among all the interacting variables. As a consequence, goal making tends to become an exercise in rhetoric, seen by many as simply a way of postponing if not avoiding hard decisions about such matters as the level of financial support for the schools, the method of allocating funds, the bases for hiring and firing teachers, the scope of services the schools are to provide, and the like.

A second reason for the confusion about goals seems to lie in the conflicting interests among and within the many different groups having a direct economic and/or political stake in the educational enterprise—parents, taxpayers, teachers, school executives, school board members, legislators, bureaucrats, commercial suppliers of plant and equipment, and, not least, the students themselves. The questions that inevitably trouble the members of these groups are: "What is there in it for me? Are the goals on which an educational assessment program is to be based consistent with my own goals? And to what extent will the program be a threat to my attainment of them?"

These are questions that must be squarely faced and coped with by educational leaders and planners if statewide assessment is to fulfill its promise. Somehow the numerous constituencies in the vast social undertaking we call education must be helped to understand that they have a *common* stake in the process, that educational assessment, when properly conceived and conducted, has the overriding purpose of increasing knowledge about what is effective in education, deepening understanding of all aspects of the educational process, opening education to all the publics concerned, and extending the ability of the schools to meet the diverse developmental needs of all students of all ages and conditions.

It is our hope that future surveys of statewide educational assessment programs will extend information on how all these problems are being dealt with so as to assist the planners-to-come in evaluating available strategies for making assessment an effective means of improving the benefits of education through informed decision making in all parts of the system.

REFERENCES

1. Tyler, R.W. "Assessing the Progress of Education." *Phi Delta Kappan*, 1965, *27*: 13-16. For a recent description of the National Assessment program in operation, see Womer, F.B. *What is National Assessment*? Ann Arbor, Mich.: National Assessment of Educational Progress, 1970.

2. Public Law 89-10, 89th Congress, H.R. 2362, April 11, 1965, Sec. 205.(a)(5).

3. Coleman, J.S. and others. *Equality of Educational Opportunity*. Washington, D.C.: U.S. Government Printing Office, 1966.

4. Educational Testing Service. *State Testing Programs: A Survey of Functions, Tests, Materials and Services*. Princeton, N.J.: Educational Testing Service, 1968.

5. Educational Testing Service. *State Educational Assessment Programs*. Princeton, N. J.: Educational Testing Service, 1971 ($4.00). The Interview Guide appears as an appendix to this full report of the survey.

6. U.S. Office of Education. *Federal/State Task Force on Educational Evaluation: An Overview*. Washington, D.C.: The Task Force, 1971.

7. Public Law 90-247, 90th Congress, H.R. 7819, January 2, 1968, Section 402.

8. Weaver, W.T. "The Delphi Forecasting Method." *Phi Delta Kappan*, 1971, *52*, 267-271.

9. Uhl, N.P. *Encouraging Convergence of Opinion, Through the Use of the Delphi Technique, in the Process of Identifying an Institution's Goals*. Princeton, N.J.: Educational Testing Service, 1971.

10. Curtis, W. H. *Project to Develop a Program Planning-Budgeting-Evaluation System Design*: First Year Final Progress Report. Chicago: Research Corporation of the Association of School Business Officials, 1968.

11. Lord, F.M. "Estimating Norms by Item-Sampling." *Educational and Psychological Measurement*, 1962, *22*, 259-267. This paper contains one of the earliest investigations of the possibilities in item-sampling for estimating the parameters of test-score distributions. This technique set the stage for what is now known as matrix sampling. See also: Lord, F.M. and Novick, M. Chapter 11, *Statistical Theories of Mental Test Scores*. Reading, Mass.: Addison-Wesley, 1968.

12. The Vermont Department of Education. *Vermont Design for Education*. Montpelier: Vermont Department of Education, 1969.

13. Bauer, R.A. (Ed.) *Social Indicators*. Cambridge, Mass.: MIT Press, 1966.

14. U.S. Department of Health, Education, and Welfare. *Toward a Social Report*. Ann Arbor, Michigan: University of Michigan Press, 1970.